Message *Distortion* in Information Cascades

ICWSM Soc2Net Workshop
June 11, 2019

Manoel Horta Ribeiro
Kristina Gligorić
Bob West

Objective: Previous studies showed conflicting results regarding the role of chocolate consumption during pregnancy and the risk of preeclampsia. We aimed to evaluate the impact of high-flavanol chocolate in a randomized clinical trial.

Study Design: […]

Results: […]

Conclusion: Compared with low-flavanol chocolate, daily intake of 30g of high-flavanol chocolate did not improve placental function, placental weight and the risk of preeclampsia. Nevertheless, the marked improvement of the pulsatility index observed in the 2 chocolate groups might suggest that chocolate effects are not solely and directly due to flavanol content.

E. Bujold et al., 2016. High-flavanol chocolate to improve placental function and to decrease the risk of preeclampsia: A double blind randomized clinical trial. *American Journal of Obstetrics & Gynecology*, *214*(1), pp.S23-S24.

Eurel

NEWS RELEASES MULTI

PUBLIC RELEASE: 1-FE

The bene
pregnanc

SOCIETY FOR MATER

sheknows

PARENTING

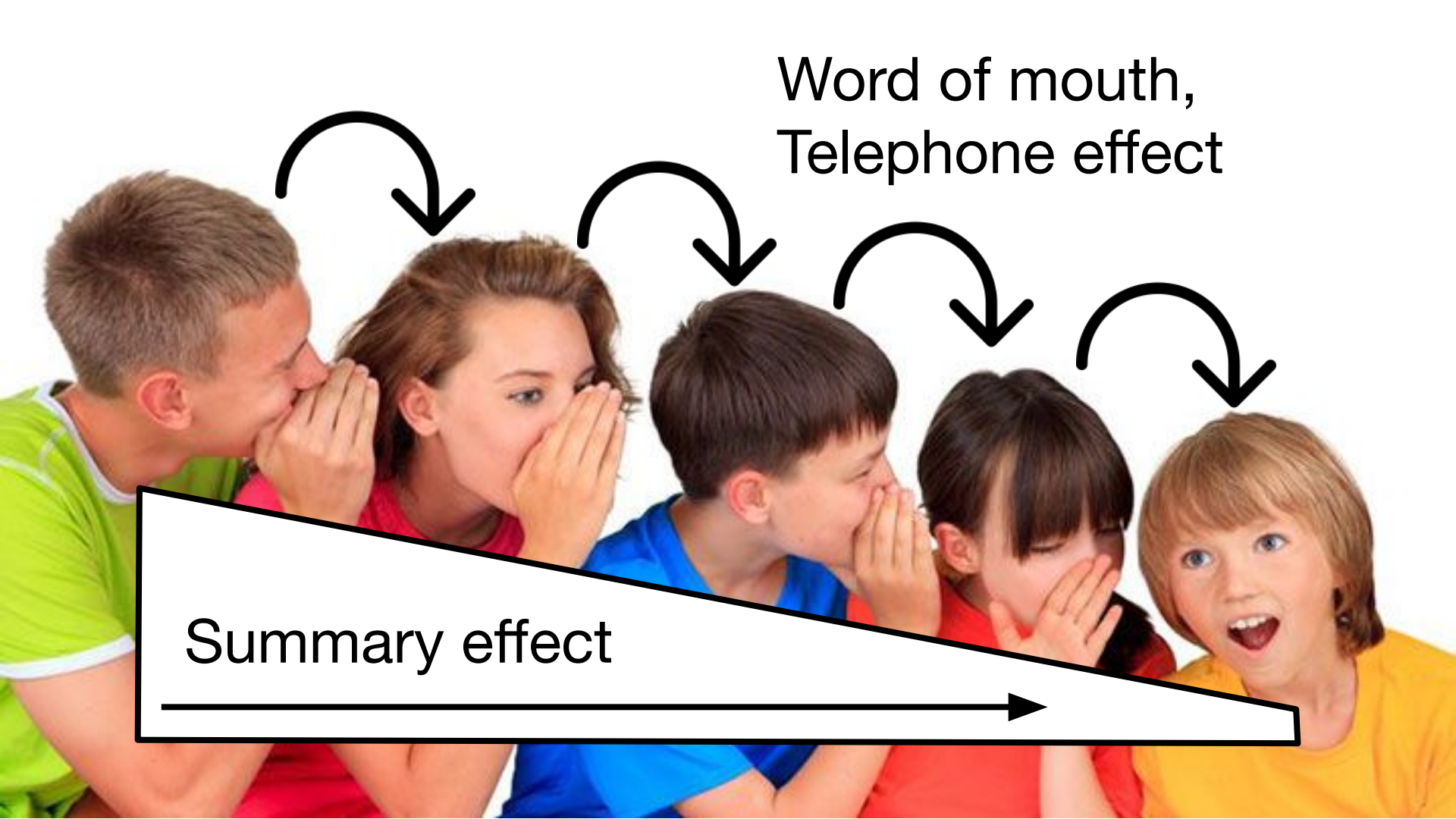
7 Reasons pregnant women need to eat a chocolate bar immediately

by Gemma Hartley

Feb 03, 2016 at 3:09 pm EDT

DAY

ring
efit fetal
t


Word of mouth,
Telephone effect
Summary effect

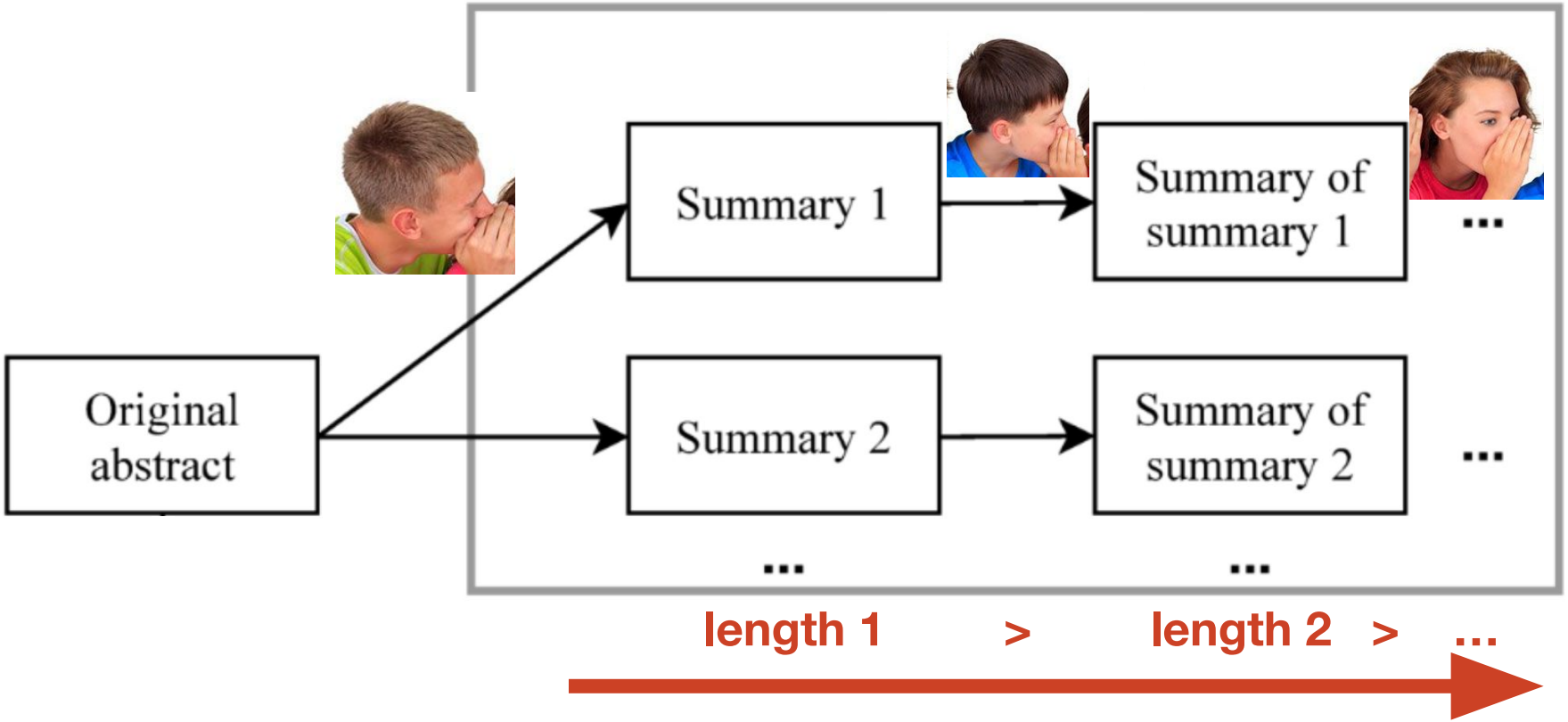
Goals of this project:

- Quantify “telephone” effect
- Tease it apart from summary effect
- Describe anatomy of “telephone” chains
- Understand how to avoid “telephone” effect


DELL
Convex Optimization
Networks
Informatik-Handbuch
KLUGE

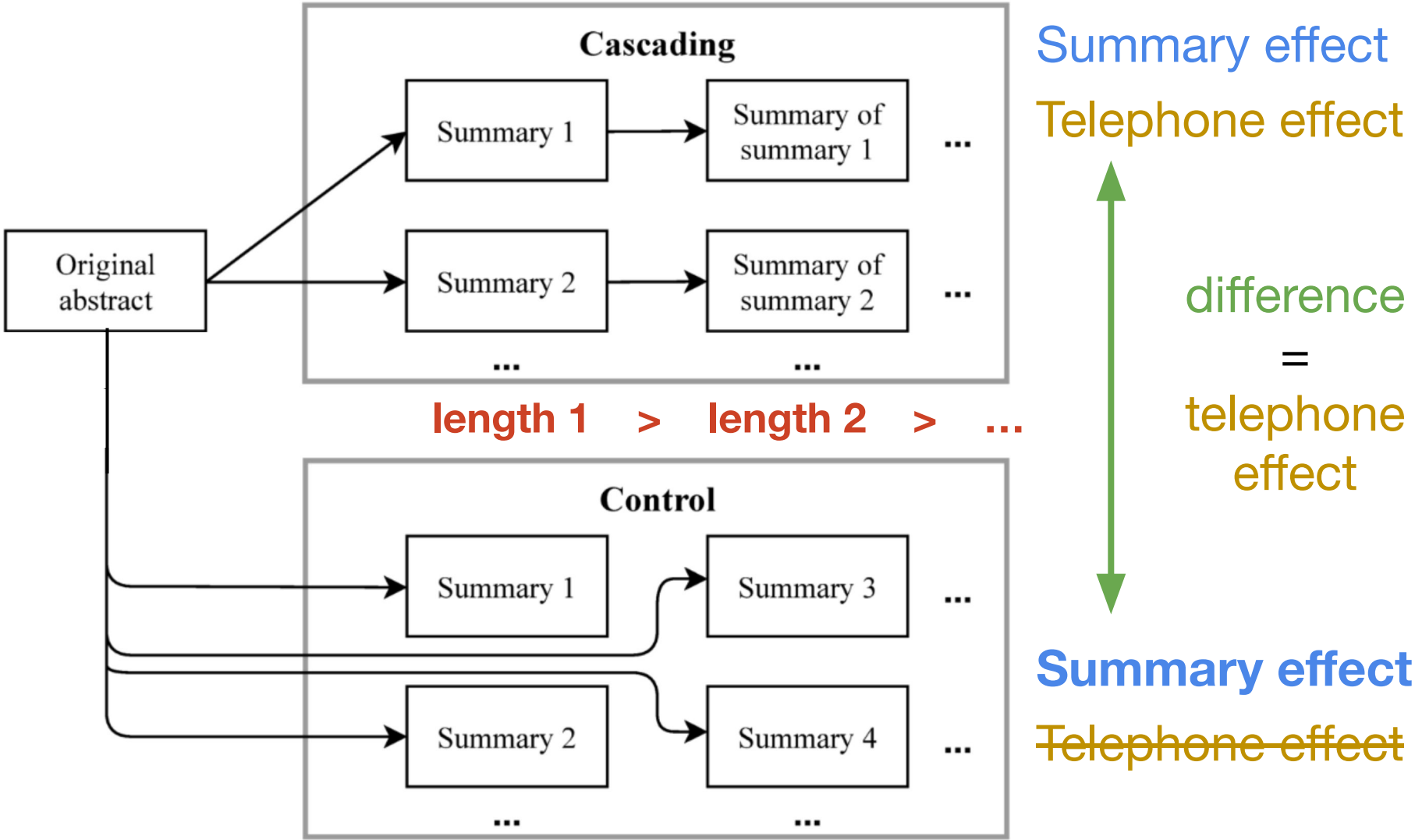

1000
800
600
400

Experiment design:
Collecting information cascades



Word of mouth → telephone effect

Decreasing length → summary effect


Cascading
Original abstract
Summary 1
Summary of summary 1
Summary 2
Summary of summary 2
length 1 > length 2 > ...
Control
Summary 1
Summary 3
Summary 2
Summary 4
Summary effect
Telephone effect
difference = telephone effect
Summary effect
~~Telephone effect~~

Dataset:
Cascades of medical information

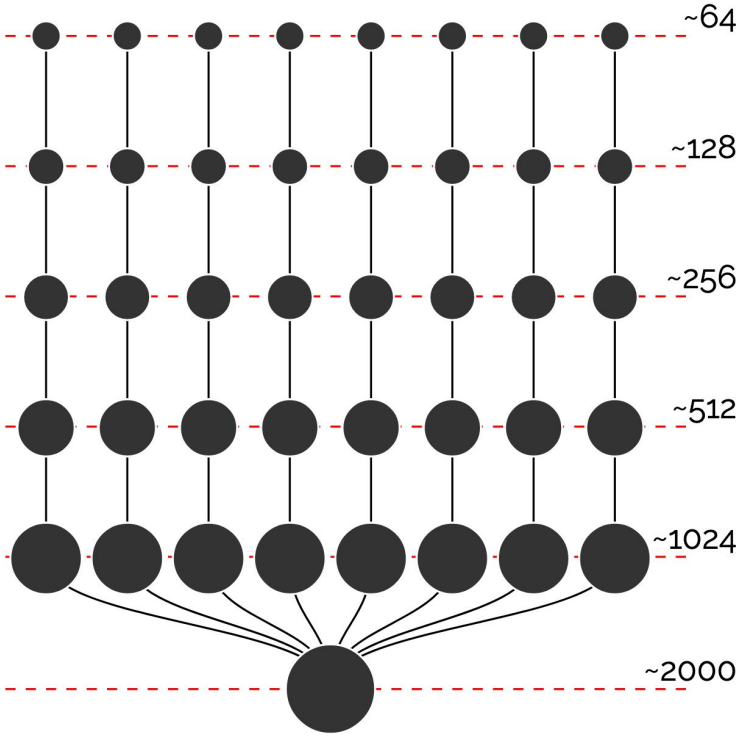
The NEW ENGLAND JOURNAL of MEDICINE

VOL. 366 NO. 2

ESTABLISHED IN 1812 JANUARY 12, 2012 NEJM.ORG

Collecting cascades via crowdsourcing

- 4 research fields of high public interest
  - Vaccination
  - Breast cancer
  - Cardiovascular disease
  - Nutrition
- 4 impactful papers (abstracts) per research field
- 8 independent cascades per abstract, collected on Amazon Mechanical Turk
  - Original abstract: ~2,000 characters
  - 5 target lengths: 1,000 > 500 > 250 > 125 > 64
- 8 control summaries per (abstract, length)
- That is, 1,280 summaries in total



Annotating and tracking information along cascades

Association of Coffee Drinking with Total and Cause-Specific Mortality

Introduction. Coffee is one of the most widely consumed beverages , but the association between coffee consumption and the risk of death remains unclear .

Methods. We examined the association of coffee drinking with subsequent total and cause-specific mortality among 229,119 men and 173,141 women in the National Institutes of Health -- AARP Diet and Health Study who were 50 to 71 years of age at baseline . Participants with cancer , heart disease , and stroke were excluded . Coffee consumption was assessed once at baseline .

Results. During 5,148,760 person-years of follow-up between 1995 and 2008 , a total of 33,731 men and 18,784 women died . In age-adjusted models , the risk of death was increased among coffee drinkers . However , coffee drinkers were also more likely to smoke , and , after adjustment for tobacco smoking status and other potential confounders , there was a significant inverse association between coffee consumption and mortality . Adjusted hazard ratios for death among men who drank coffee as compared with those who did not were as follows : 0.99 (95 % confider ng less than 1 cup per day , 0.94 (95 % CI , 0.86 to 0.93) for 2 or 3 cups , ups , and 0.90 (95 % CI , 0.85 to 0.96) for 6 or more cups of coffee per day (P <= 0.001 for trend) ; the respective hazard ratios among women were 1.01 (95 % CI , 0.96 to 1.07) , 0.95 (95 % CI , 0.90 to 1.01) , 0.87 (95 % CI , 0.83 to 0.92) , 0.84 (95 % CI , 0.79 to 0.90) , and 0.85 (95 % CI , 0.78 to 0.93) (P < 0.001 for trend) . Inverse associations were observed for deaths due to heart disease , respiratory disease , stroke , injuries and accidents , diabetes , and infections , but not for deaths due to cancer . Results were similar in subgroups , including persons who had never smoked and persons who reported very good to excellent health at baseline .

Conclusion. In this large prospective study , coffee consumption was inversely associated with total and cause-specific mortality . Whether this was a causal or associational finding can not be determined from our data .

"Keyphrases"

Coarse category	Fine category	
Participants	Sex	and
	Age	71 years
	Condition	se and
	Location	
	Sample size	1
Intervention	General	ffee
	Duration	etween
	Intensity	
	Control	
Outcomes	General	The study measured long-term total and cause-specific mortality.
	Effect strength	N/A
	Adverse effects	N/A
Conclusion	General	Coffee consumption was inversely associated with total and cause-specific mortality in non-smokers.

Fact about Participants/Sex:
“The study was performed in women and men.”

Summary:
“A study of coffee drinking and mortality initially was positive. Results were reversed when it was found that smoking was also a factor.”

“Fact scores”:

A	The fact is entirely captured in the text, omitting only insignificant details.
B	The essence of the fact is captured in the text, but a significant amount of detail was omitted.
C	The fact is not, or only insufficiently, captured in the text.
D	The fact contradicts the original text.

Example cascade

A: fact fully captured
B: fact partially captured
C: fact missing
D: fact contradicted

A study on the association of drinking coffee with the risk of death remains unclear. There was an examination of the association of drinking coffee with total and cause-specific mortality among 229,119 men and 173,141 women in the NIH-AARP Diet and Health Study who were 50-71 years of age at baseline. Participants with cancer, heart disease, and stroke were excluded. Consumption was assessed once at baseline. During follow-up between 1995 and 2008, a total of 33,731 men and 18,784 women died. In age-adjusted models, the risk of death was increased among coffee drinkers. However, coffee drinkers were more likely to smoke. After adjustment for tobacco-smoking status and other potential confounders, there was a significant inverse association between coffee consumption and mortality. Results were similar in subgroups, including persons who had never smoked and persons who reported very good to excellent health at baseline. In this large prospective study, coffee consumption was inversely associated with total and cause-specific mortality.

Participants					Intervention				Outcomes			Conclusion
Sex	Age	Condition	Location	Sample size	General	Duration	Intensity	Control	General	Effect strength	Adverse effects	General
A	A	A		A	A	A			A			A
men; women	50; 71	cancer; heart diseases; stroke		229,119; 173,141	coffee	1995; 2008			total; cause-specific; mortality			

...

A study of coffee drinking and mortality initially was positive. Results were reversed when it was found that smoking was also a factor.

C	C	C		C	A	C			B		A
					coffee				mortality		

Coffee can lead to earlier death, but smoking is also a factor.

C	C	C		C	A	C			B		D
					coffee						

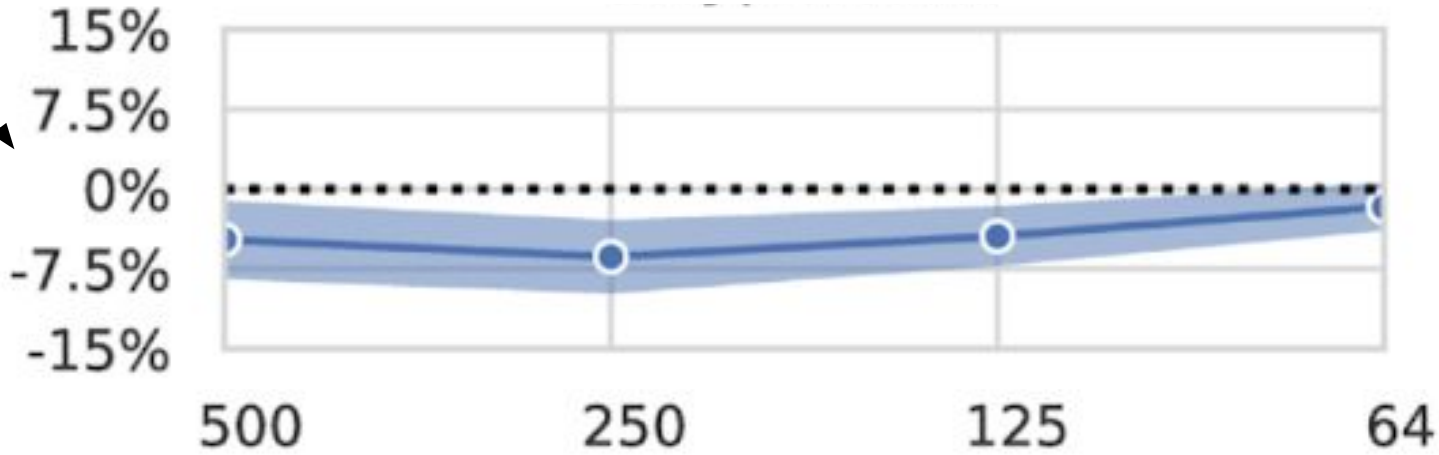
Research questions

- **RQ1:** How strong is the telephone effect?
- **RQ2:** How does info persist hop by hop?
- **RQ3:** Should I be extractive or abstractive?

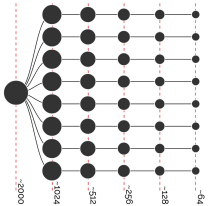
RQ1
How strong is the telephone effect?

Keyphrase persistence

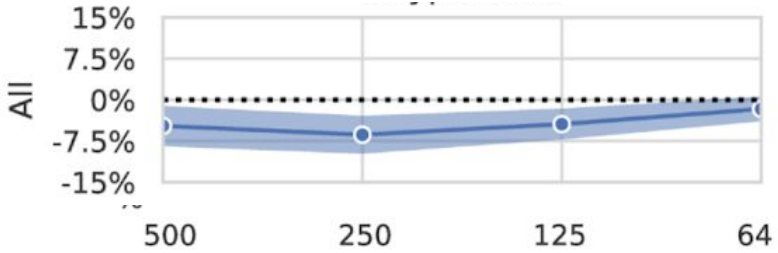
Difference (cascades minus control) of fraction of summaries in which keyphrase is present

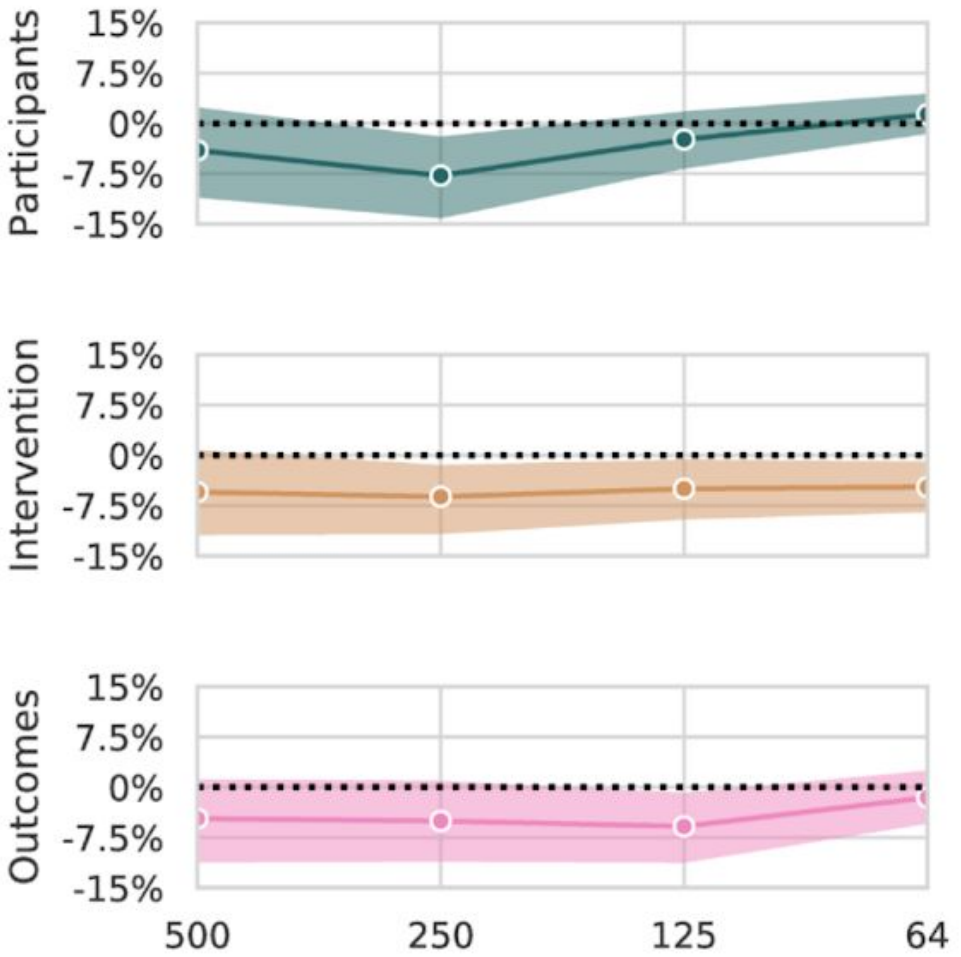


Target length

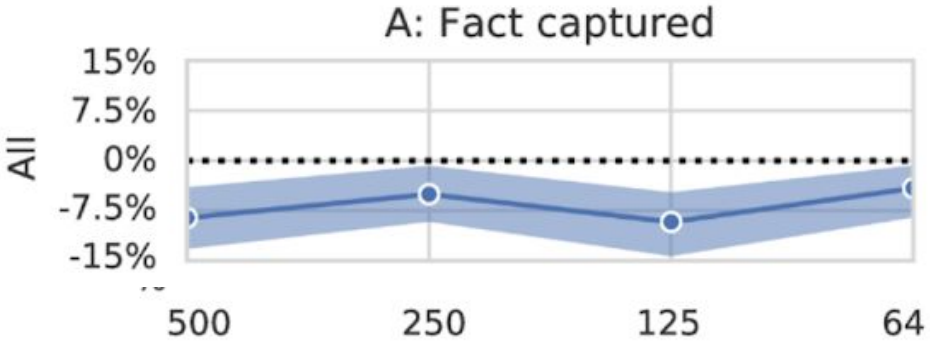


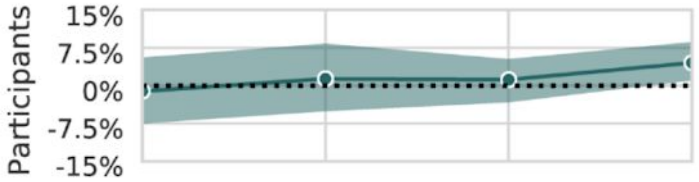
Keyphrase persistence

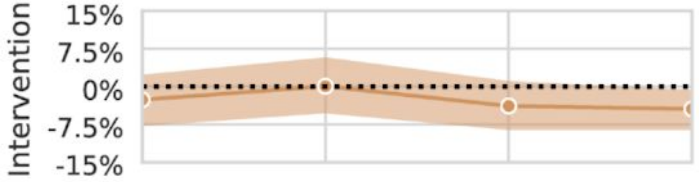


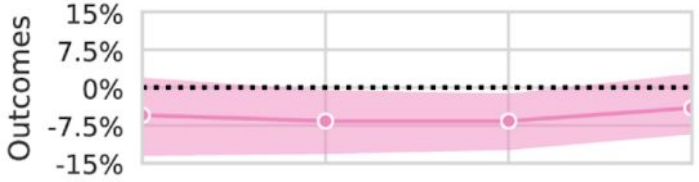


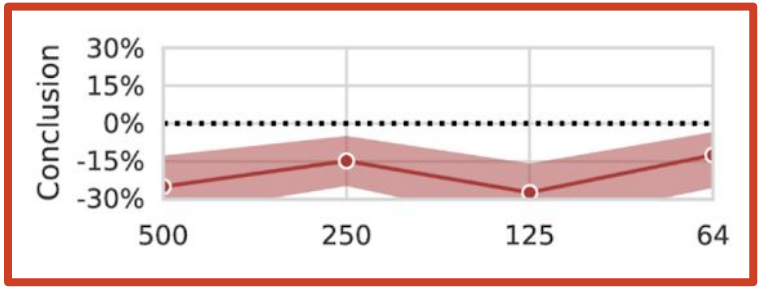
Fact persistence





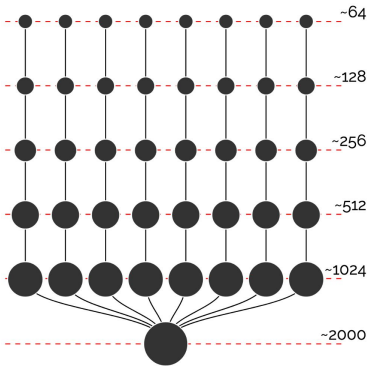


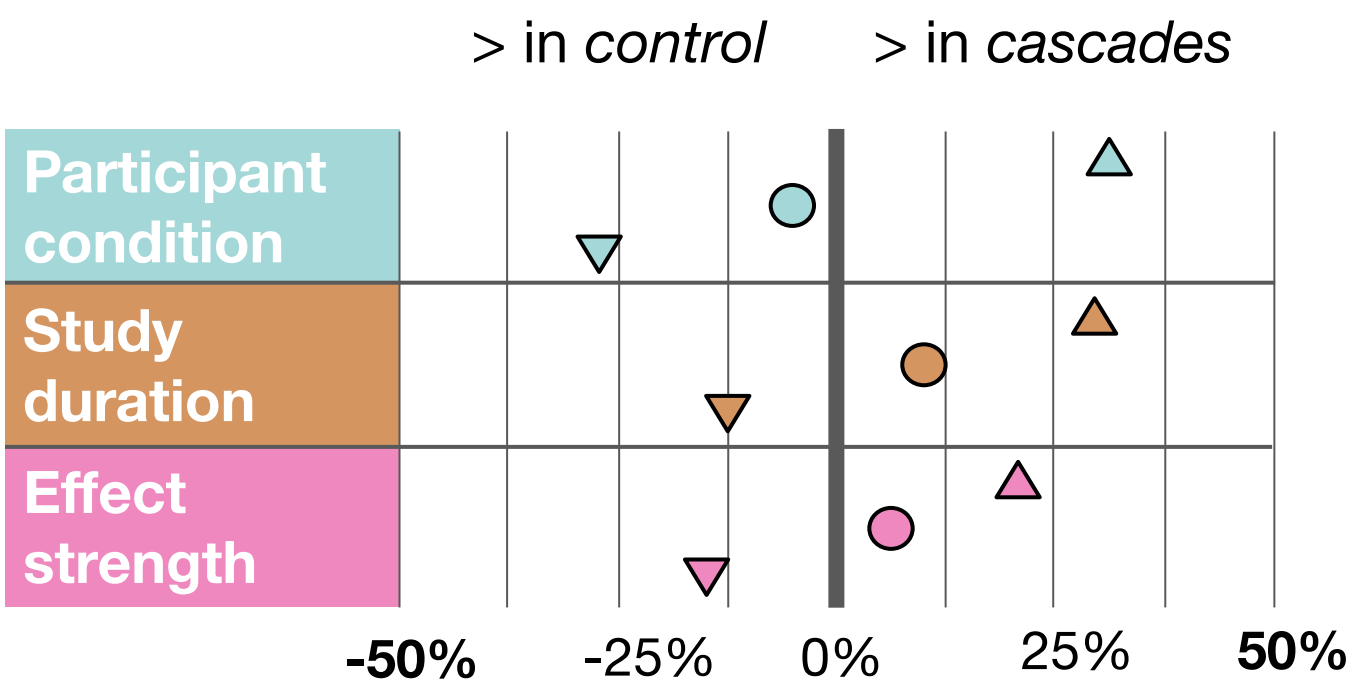




Is the telephone effect sometimes useful?

Is the telephone effect sometimes useful?

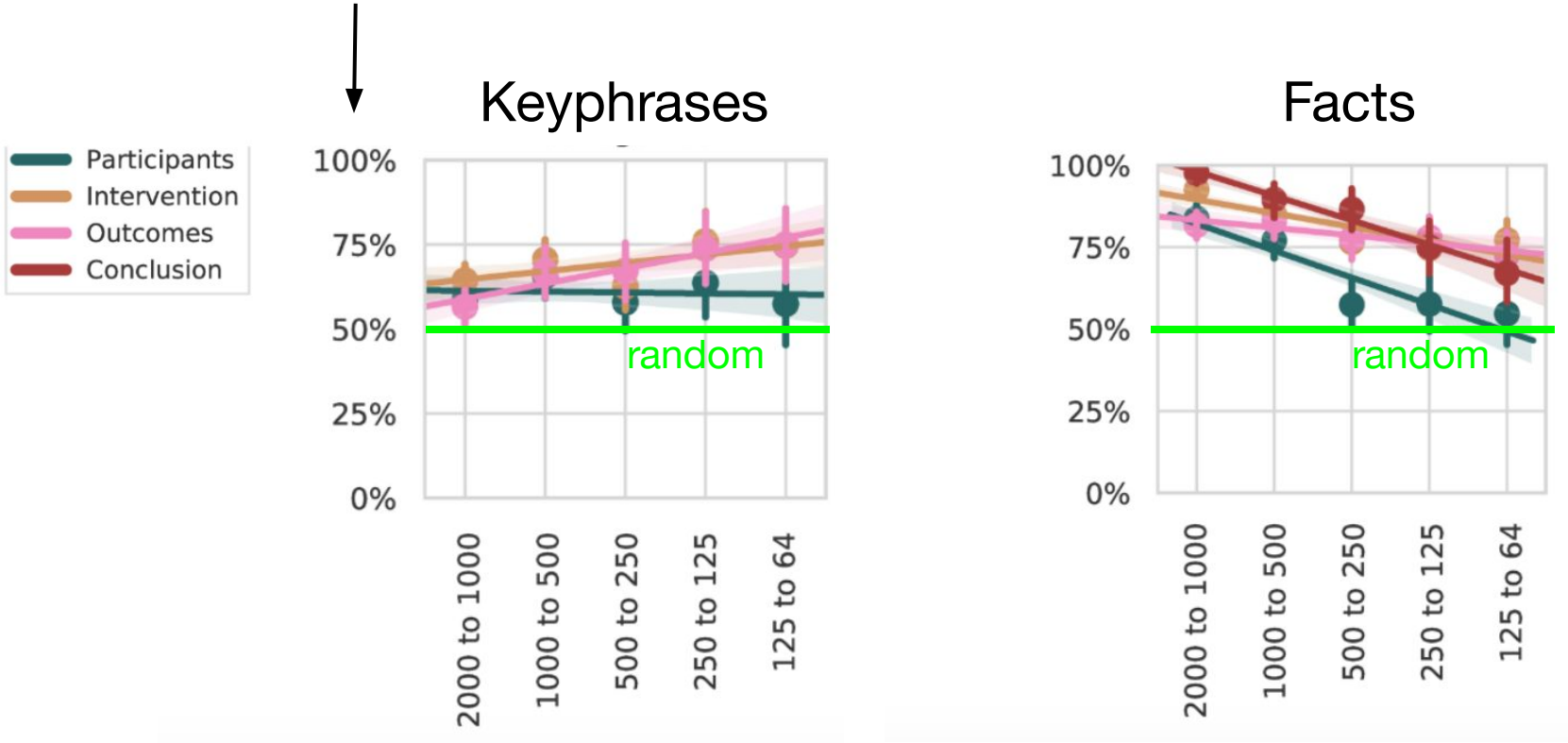




Difference in % fully preserved facts (cascades minus control), averaged over all target lengths

RQ2
How does info persist hop by hop?

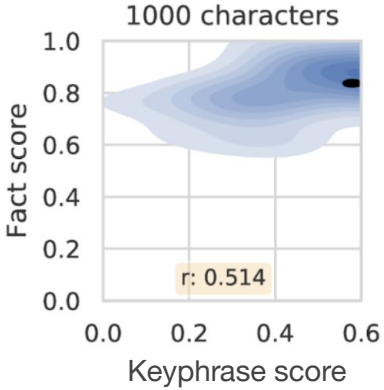
Given that a keyphrase has already survived *k* hops, how likely is it to survive one more?

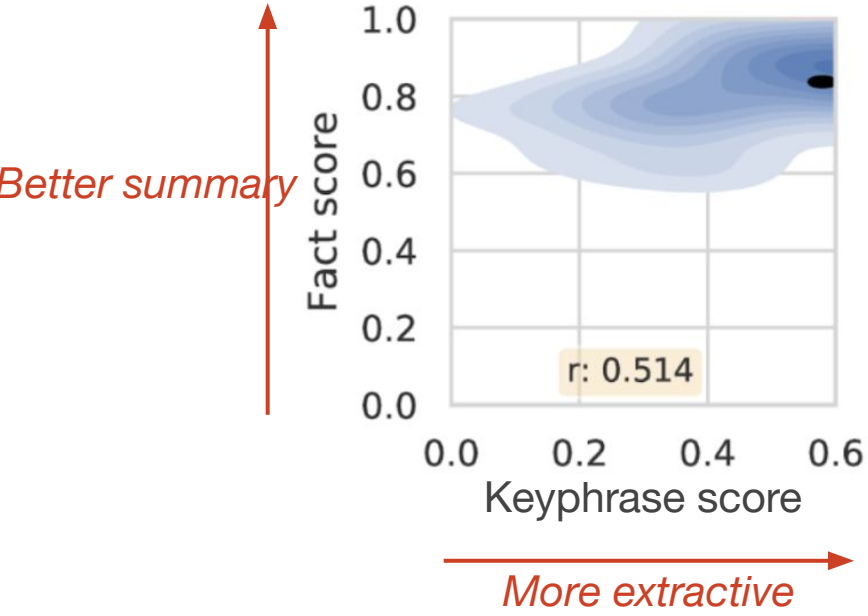


RQ3

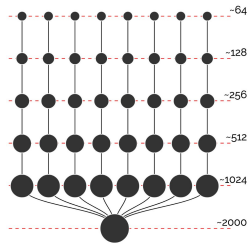
Should I be extractive or abstractive?


Extractive summary:
Four score and seven years ago our fathers brought forth a new nation dedicated to liberty and equality.
Abstractive summary:
87 years ago, 'Murica was founded, a country of free and equal citizens. U-S-A, U-S-A, U-S-A!

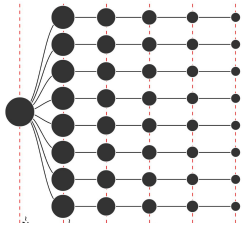




- Fix quality (fact score) of source summaries *S*
- Compare summaries of extractive *S* vs. abstractive *S*
- Result: quality (fact score) of summaries of extractive *S* is higher



Summary



The NEW ENGLAND JOURNAL of MEDICINE

VOL. 366 NO. 2

A study on the association of drinking coffee with the risk of death remains unclear. There was an examination of the association of drinking coffee with total and cause-specific mortality among 229,119 men and 173,141 women in the NIH-AARP Diet and Health Study who were 50-71 years of age at baseline. Participants with cancer, heart disease, and stroke were excluded. Consumption was assessed once at baseline. During follow-up between 1995 and 2008, a total of 33,731 men and 18,784 women died. In age-adjusted models, the risk of death was increased among coffee drinkers. However, coffee drinkers were more likely to smoke. After adjustment for tobacco-smoking status and other potential confounders, there was a significant inverse association between coffee consumption and mortality. Results were similar in subgroups, including persons who had never smoked and persons who reported very good to excellent health at baseline. In this large prospective study, coffee consumption was inversely associated with total and cause-specific mortality.

- Question: How is info distorted as it is passed on by word of mouth?
- Experimental design: experimental study on crowdsourcing platform
- Study performed: propagation of info from medical abstracts
- Careful manual coding of keyphrases and facts in all abstracts and summaries

RQ 1: How strong is the telephone effect?

- Strong! Much more info lost in cascades vs. controls
- Especially bad for most important info (conclusions of papers)
- If source summary was good, telephone effect is useful!

RQ 2: How does info persist hop by hop?

- Surviving keyphrases ever *more* likely to survive further
- Surviving facts ever *less* likely to survive further

RQ3: Should I be extractive or abstractive?

- Extractive!

Dataset available:
https://go.epfl.ch/distortion

(Demo)


SO WHAT?

- Messages distorted w/o malicious actors
- Medical abstracts: most important info most prone to distortion
- Solution angles:
  - Be extractive! Keep catchy keyphrases!
  - Show multiple summaries

Future work should

- move from the lab to the wild:
  - real cascades on real platforms
- study more settings:
  - news,
  - political opinions and statements
- build models of message distortion


Thanks!
Questions?
robert.west@epfl.ch